植物分类学报 42(5):464-479(2004)
Acta Phytotaxonomica Sinica

·综述·

# 多年生植物模式物种基因组研究的历史及进展

[1]尹佟明* [2]朱其慧 [1]黄敏仁 [1]王明庥

[1](南京林业大学森林遗传与基因工程开放研究实验室 南京 210037)

[2](中国科学院植物研究所系统与进化植物学重点实验室 北京 100093)

# History and progress of the genomics studies in the model system of perennial plant species

[1]YIN Tong-Ming* [2]ZHU Qi-Hui [1]HUANG Min-Ren [1]WANG Ming-Xiu

[1](*The Key Laboratory of Forest Genetics and Gene Engineering*, *Nanjing Forestry University*, Nanjing 210037, China)

[2](*Laboratory of Systematic and Evolutionary Botany*, *Institute of Botany*, *the Chinese Academy of Sciences*, Beijing 100093, China)

**Abstract** Strong justifications have been developed for why woody plants should be viewed as model systems in plant biology. The genus *Populus* possesses many characteristics that are conductive to functional genomic studies, and therefore leads to its emergence as a model system in extrapolating findings in perennial plant species that are different from annual herbaceous plants. With the proceeding of the whole genome sequencing, poplars will be act as a wide reference for functional genomics studies in perennial plant species and will also contribute towards answering some fundamental scientific questions. This paper reviewed the history and progress of the poplar genome studies and the potential keen topics in the future. The contents mainly address on: (1) the somatic genetics studies in *Populus*; (2) the genomics studies carried out in *Populus*, including genetic mapping, genome sequencing, physical map construction, microarray analysis and linkage disequilibrium analysis; (3) the potential application of the genome information of *Populus* for facilitating our understanding of some basic scientific questions.

**Key words** Genome, perennial plant, model system.

**摘要** 木本植物有许多不同于一年生草本植物的生物学特性,生物学家提出将木本植物作为研究多年生植物的模式体系。杨属 *Populus* 树种由于研究基础较好且基因组较小,目前已被广泛地接受作为多年生植物基因组研究的模式物种。随着杨属树种全基因组序列的测定,杨属树种在多年生植物的功能基因组研究及一些基础科学问题的研究中将发挥重要作用。本文综述了杨属树种基因组研究的历史、进展及将来的研究热点,旨在为我国多年生植物基因组研究提供参考和借鉴。本文主要论述了以下几个方面的内容:(1)对杨属树种开展的细胞遗传学研究;(2)在分子水平上对杨属树种进行的基因组研究,内容包括遗传作图、基因组测序、物理图谱构建、基因芯片及连锁不平衡分析;(3)杨属树种基因组信息在探讨一些基础科学问题中的潜在应用。

**关键词** 基因组;多年生植物;模式物种

森林是最重要的陆地生态系统并且是人类最主要的可持续利用的资源,但是我们对木本植物的认识相对于许多一年生植物和重要作物而言还比较贫乏,因此为有效地经营现有的森林资源,需要对决定树木适应性及生产力等的遗传机制进行更进一步的了解。

---

2004-02-25 收稿,2004-07-16 收修改稿。

基金项目:国家自然科学基金(30200224,30230300);霍英东基金(81024)资助(Supported by the National Natural Science Foundation of China (Grant Nos. 30200224, 30230300) and Huo Yingdong Education Foundation (Grant No. 81024))。

* 通讯作者(Author for correspondence. E-mail: <yintm@njfu.edu.cn>)。

生物学家对于为何需要将木本植物视为一种特殊的生物模式进行了许多论述（Bradshaw et al.，2000；Taylor，2002）。他们认为：与一年生植物相比，木本植物大多为多年生植物，一般个体高大且具有复杂的树冠结构，木本植物具有次生木质部，木本植物随季节变换有休眠特性，同时在生长过程中有幼龄期到成年期的转换。目前在基因组研究方面较为深入的植物，大多是一年生草本植物，对于多年生植物而言，还没有建立一套合适的生物学研究体系。杨属 *Populus* L.树种由于其特殊的生物学特性，目前已被广泛地接受作为多年生植物基因组研究的模式物种（Bradshaw et al.，2000；Brunner et al.，2004；Taylor，2002；Wullschleger et al.，2002a）。美国能源部基因组联合研究所对毛果杨 *P. trichocarpa* Hook.无性系'Nisqually-1'进行了全基因组序列测定：迄今，利用鸟枪法（shotgun）已完成了约为基因组全长 8 倍的基因组序列测定。杨属树种是第一个测定全基因组序列的多年生植物并且是第三个开展全基因组测序的植物（Brunner et al.，2004；Wullschleger et al.，2002a）。随着全基因组序列的测定，杨属树种在多年生植物的功能基因组研究中将起到越来越重要的作用。

## 1　杨属树种基因组细胞遗传学研究

对杨属树种基因组研究的历史最早可追溯到 1921 年 Graf 的工作（Graf，1921）。他发现加杨 *P. canadenis* Moench 和欧洲山杨 *P. tremula* L.在已经过减数分裂的胚囊母细胞中共有 4 条染色体。这一结果很快被发现是错误的。到 1924 年，Blackburn 和 Harrison 根据对 7 个杨属树种和 17 个柳属 *Salix* L.树种的细胞学观察确定了杨柳科 Salicaceae 植物经减数分裂后的细胞具有 19 条染色体（Blackburn，1924）。在柳属树种中，二倍体、四倍体和六倍体均有发现。自 1924 年至今对杨属树种所进行的细胞学研究发现，杨属 5 个组（section）的所有树种主要以二倍体（2n = 38）的形式存在，但是在白杨组 sect. *Leuce* Duby（= sect. *Populus*）树种中也发现了少数几个种存在三倍体形式（Smith，1943；Harrison，1924；Johnson，1940）。杨属树种的染色体很小且许多染色体大小相差很大。Blackburn 和 Harrison 于 1924 年最先对欧洲山杨染色体大小的变异进行了研究，发现："染色体间大小变异很大，其中 9 条染色体的大小相近；另外有 9 条比第一组的 9 条要大一些并形成大小呈梯度变化的另一组染色体；最后还有 1 条染色体的大小约为其他任一条染色体大小的二倍左右。"其后的许多研究在不同的杨属树种中都观察到了这一条'硕大'的染色体（"giant chromosome"）（Muntzing，1936；Smith，1943）。基于形态学的观察，Van Dillewijn（1940）对染色体的二级相关（secondary association）进行了研究，他把杨属树种染色体划分成了 3 组各具有 3 条相近大小的染色体，4 组各具有 2 条相近大小的染色体和 1 组只含 1 条'硕大'染色体的组；另外，还有一条较小的染色体与'硕大'染色体相对应。因而提出了杨属树种的染色体可能起源于 8 条原始染色体的假说。根据他的假说，8 条原始染色体中，其中的 4 条发生了 3 次复制，其他的 4 条原始染色体发生了 2 次复制，最后在 1 组 3 次复制的染色体组中有两条染色体融合，从而形成了 1 条硕大的染色体。在其后的细胞学研究中发现，在减数分裂联会过程中，偶尔出现错配的三价体和四价体。这种错误联会的发生，可能是由于某些染色体间的相似性所导致。其后的研究虽然不很支持 Van Dillewijn 的详细划分，但都观察到某些染色体大小相近（Smith，1943）。然而仅根据染色体形态的

研究，推测染色体的进化和得出确定的原始染色体数目是不可靠的。

在已研究的杨属树种中均观察到一条‘硕大’染色体（“giant chromosome”）的存在，按当时的研究认为杨树可能存在性染色体对，即‘硕大’染色体为X染色体，而较小的染色体中可能有一条为Y染色体。通过减数分裂过程中染色体联会的细胞学观察，迄今没有发现稳定的形态相差很大的染色体对（chromosome pair）可以被解释为性染色体，从而否定了杨属树种中特化的性染色体对的存在（Smith，1943）。

早期的杨属树种细胞遗传学研究主要着眼于探讨杨属树种的进化及种分化的遗传学机制。一般杨属树种在系统上可分为黑杨组sect. *Aigeiros* Duby、青杨组sect. *Tacamahaca* Spach、大叶杨组sect. *Leucoides* Spach、白杨组sect. *Populus* 和胡杨组sect. *Turanga* Bge.（王战等，1984）。对于种分化的遗传机制，目前有两种主要的假说模型，一种是以Mayr（1970）为代表提出的地理隔离模型（allopatric model），这种模型认为在地理上有隔离物种，由于遗传变异在地理隔离的群体中积累，最终导致了群体间的生殖障碍，从而导致了物种分化。另一种是以White（1978）为代表提出的染色体结构差异导致生殖隔离的模型（stasipatric model），这种模型认为由于染色体重排产生了染色体结构上的差异进而导致生殖上的隔离。后来有许多的新模型提出，但基本上都是对于上述两种模型的改进（Gibson，1984）。近几十年来的大量比较基因组研究，尤其是在哺乳动物和禾本科植物的研究中发现了许多种间染色体重排的证据（Gibson，1984；Devos & Gale，1997；Moore et al.，1995）。由染色体重排而导致的染色体结构上的差异越来越成为一种被广泛接受的解释物种分化遗传机制的假说。然而，在以往研究中所检测到基因组重排到底是物种分化的原因还是结果，一直是一个无法回答的问题。这主要是因为研究所涉及的物种的生殖隔离已经形成，因而无法确定分化过程中的原始机制。杨属树种可能为这一问题的研究提供一种较为理想的模式，杨属为单系起源，组间分化远近不同。对青杨组和黑杨组种间杂种的细胞遗传学研究发现，组间和组内杂种在花粉不育性（pollen sterility）、单价体形成（univalent formation）、倒位桥频率（inversion bridge frequency）方面均无明显差异，这些参数在同一组内有地理隔离和无地理隔离的种间也没有明显差别。早期的细胞遗传学研究认为，杨属树种的分化主要是由于地理、适应性及生理上的隔离，遗传和染色体结构上的差异可能是一种次要的因素（Smith，1943）。一些杨属树种如毛果杨和美洲黑杨*P. deltoides* Marsh.的天然分布域存在重叠并有广泛的天然基因渐渗，因此非遗传因素（地理、适应性及生理因素）导致的隔离并不能很好地解释杨属树种的分化，虽然在杨属树种杂种的细胞遗传学研究中发现了大量的染色体结构差异的证据，但是许多种的天然杂交障碍都可以通过人工杂交克服，所以染色体结构的差异并不一定会导致生殖上的隔离，基于染色体重排的假说也不能很好地解释杨属树种的分化。如果可以在杨属树种不同组间、种间建立起动态的基因组重排随遗传距离变化的过程，将能够加深我们对于种分化遗传机制的认识。由于杨属树种的染色体很小，其后发展起来的染色体分带技术在杨属树种染色体研究的应用中受到很大限制，因此没有能够在杨属树种的细胞遗传学研究中得到应用。杨属树种细胞遗传学的研究自20世纪40年代后基本没有新的进展。但这些早期的经典研究，即使今天来看仍具有重要的学术价值。

对于上面提到的问题，仅基于经典细胞遗传学的研究结果难以做出确定的回答。同

时由于杨属树种是一种重要的工业用材树种，也需要一种更为有效的技术可应用于杨属树种的遗传改良中。随着分子标记辅助选择在重要作物及畜牧改良中的应用，这一技术也有希望突破木本植物长育种周期的瓶颈。自20世纪90年代开始，在分子水平上进行的基因组研究在木本植物中开始广泛开展。

## 2 分子水平上的杨属树种基因组研究

### 2.1 遗传图谱

最早在杨属树种基因组中进行的分子标记间的连锁分析是利用同工酶和RFLP（restriction fragment length polymorphism，限制性长度多态性）标记进行的（Liu & Furnier，1993），真正意义上的杨属树种的第一张分子标记连锁图谱由华盛顿大学构建，这张图谱主要由RAPD（random amplified polymorphic DNA，随机扩增DNA多态性分析）标记组成，另外包含了少量的RFLP和STS（sequence-tagged site，序列标记位点）标记，图谱覆盖的基因组长度约为全基因组的50%（Bradshaw et al.，1994）。其后，在杨属树种中先后报道了几张由RAPD和AFLP（amplified fragment length polymorphism，扩增性片段长度多态性）标记构建的杨属树种图谱（Wu et al.，2000；Zhang et al.，2004；Yin et al.，2001；Yin et al.，2002）；2001年，Cevera等报道了4张共含99个SSR（simple sequence repeats，串联的简单重复序列，又称微卫星）标记的图谱（Cevera et al.，2001）。第一张连锁群数目与杨属树种单倍体染色体数目对应的图谱由美国橡树岭国家实验室构建（Yin et al.，2004b），该图谱覆盖了杨属树种的全基因组，根据图谱显示的单倍体细胞交叉数（chiasmata）、基于标记间距的累加长度及1000次位点随机抽样进行的估计，3种方法获得的基因组长度均为2400 cM（左右）。杨属树种图谱构建的发展过程基本上代表了木本植物图谱构建的发展过程。统观近10年来构建的木本植物遗传图谱主要存在以下几个方面的问题：(1) 图谱标记主要以随机显性标记为主。利用随机显性标记虽然可以快速建成遗传图谱，但无法进行图谱间的比较，建成的图谱及利用图谱获得的QTL（quantitative trait locus，数量性状位点）信息都具有组合特异性，极大地限制了遗传图谱的应用价值。同时显性标记不能区别杂合等位基因，而木本植物多为基因位点杂合度较高且连锁相（linkage phase）未知的远交物种（outbred species），对特定的杂交组合而言，每个基因位点可能最多具有4个分离的等位基因。木本植物中最常用的对于显性标记所采取的拟测交分析技术（pseudotest cross）是将不可见标记（invisible alleles）作为纯合的哑等位基因（null allele）处理（Grattapaglia & Sederoff，1994），因此QTL的分析存在较大误差。(2)作图个体数量有限。Zamir（1986）指出，衡量图谱质量的标准并不是以标记数目为准，一定的作图群体大小只能建成一定饱合度的图谱，如人类具有5840个微卫星标记的遗传图谱中，由于作图群体大小的限制，图谱位置确定的标记数只有970个（Remington et al.，1999）。已构建的木本植物遗传图谱，作图个体数大都在100个以内（Temesgen et al.，2001；Barreneche et al.，1998；Nelson et al.，1994；Binelli & Bucci，1994；Bradshaw et al.，1994；Cevera et al.，2001；Costa et al.，2000；Devey et al.，1996；Echt & Nelson，1997；Grattapaglia & Sederoff，1994；Mukai et al.，1995；Nelson et al.，1993；Remington et al.，1999；Travis et al.，1998；Viruel et al.，1995）。如果作图群体大小一定，单纯增加标记数目，并不能对提高QTL分析效率提供新的信息。对QTL分

析的计算机模拟分析显示，作图个体数在 400 个以上时，检测到的 QTL 效应才趋于稳定（Kearsey & Farquhar，1998）。（3）基因型错误。基因型错误在木本植物遗传图谱构建中是一个较为严重的问题，由于标记本身的不可比性，图谱间无法比较。在已构建的木本植物图谱中，大多数图谱揭示的基因组长度随标记的增加呈非收敛性增加，而且随着标记数的增加往往非连锁标记的数量也在增加，这与标记基因型的错误有关。根据定义，基因组的遗传长度应该等于单倍体细胞中观察到的期望交叉数的 100 倍，例如杨属树种二倍体细胞中，如果观察到的期望交叉数为 48.2 个，则基因组的遗传长度应为 2410 cM。当图谱上的标记覆盖全基因组后，这一数值随标记数的增加是收敛的，不会随标记数的增加而产生变动。在人类及畜牧作图中，由于采用多个家系，标记在不同的家系内的重组频率不一致（recombination frequency heterogeneity），从而导致一些标记无法定位而只能作为辅助标记（accessory marker）处理。这一概念在木本植物作图中也被借用，但木本植物的图谱都是基于单一家系，随着标记数目的增加，许多标记会定位于同一图位（bin），确定这些标记顺序需要检测到新的重组个体，因而需要更大的作图群体，但不会引起标记定位混乱的问题。所以无法定位的辅助标记的产生是因为标记的基因型错误所致，而不能归因于双交换等因素。根据遗传长度的定义，我们可知对于一个细胞而言，每条染色体上实际发生的交叉数一般在 1–2 次，所以距离较近的双交换基本上是由基因型错误导致。计算机模拟显示，3%的基因型错误会导致遗传图谱的长度加倍，而 5%的错误，会使大多数 QTL 的效应检测不到（Abecasis et al.，2001；Kearsey & Farquhar，1998；Yin et al.，2003）。而在以往的研究中，这样比例的错误是常见的（Tsuchiya，1984），这在木本植物图谱构建的过程中是一个值得引起注意的问题。

以往构建的大多数木本植物遗传图谱，所获得的信息是零散的。另外，遗传长度与物理长度在不同染色体区域对应的长度不同。根据上述对遗传长度的介绍，每 cM 遗传长度在松属树种 *Pinus* L. spp. 中对应的平均物理长度为 500–1000 万碱基对（Echt & Nelson，1997；Yin et al.，2003），而在杨属树种中也有大约 22 万碱基对左右（Yin et al.，2004b），所以基于家系分析所获得的距目的基因较近的标记，它们间的物理距离还是很大的。一般认为，木本植物连锁不平衡程度很低，在天然群体中，即使间距几千碱基对的距离，标记与目的基因间的相关（association）也可能丧失，因而不能确定它们在群体其他个体中的连锁相，这样就有可能不能在其他家系的选择中应用（Strauss et al.，1992）。近年来，木本植物遗传图谱的构建工作主要致力于开发高度保守的标记并进行比较遗传图谱构建，迄今在部分树种中已发表了一些含有一定数量保守标记（如 SSR 和 EST：expressed sequence tags，表达序列标签等）的图谱（Cevera et al.，2001；Barreneche et al.，1998；Temesgen et al.，2001），但保守标记的数量较少，不能满足基因组详细比较的需要。

比较基因组在许多物种研究中显示（Devos & Gale，1997；Whitkus et al.，1992；Paterson et al.，2000；Anh & Tanksley，1993；Lagercrantz，1998），标记和 QTL 在一定的分类水平上存在着广泛的同线性顺序（colinearity），这种基因组相似性的存在，可以有助于在不同的物种中发现不同的有用基因、等位基因及 QTL（Yin et al.，2002）。同时通过基因组比较，可以在不同物种中清晰识别垂直位点（orthologs）和水平位点（paralogs），从而在一个物种中建立的图谱可以代表有亲缘关系的不同物种的图谱。这样可以将在一定分类等级内的物

种在图谱的基础上作为一个大遗传系统加以研究（Brunner et al.，2004）。通过比较基因组研究，可以解决前面提到的许多问题。由于基因组序列组装的需要，美国橡树岭国家实验室开展了大规模的 SSR 标记作图研究，根据利用最短重叠群法（minimum tilling path）组装的 BAC 文库的末端序列设计了近 4000 个 SSR 引物，并建成了一张 SSR 标记位点间距约为 3.8 cM 的微卫星图谱（Yin et al.，未发表）。微卫星标记的引物结合序列（sequence of the priming site）在杨属不同种间存在高度的保守性，根据 Tuskan 等（2004）的研究，SSR 在杨属树种不同组间有很高的扩增成功率，如在青杨组、大叶杨组及黑杨组中为 80% – 99%；在白杨组中为 70% – 80%；在胡杨组中为 70%左右；即使在柳属树种中，通用率仍在 30% – 50%（Tuskan et al.，2004）之间。通过与 Eckenwalder 等（1996）研究的杨属树种不同种间的遗传距离的比对，微卫星引物（根据毛果杨序列设计）在不同种间的通用率和他们相对于毛果杨的遗传距离存在极显著相关。目前已有大量研究显示 SSR 标记在自然群体中揭示的变异频率是最高的（Altukhov & Salmenkova，2002；Byrne et al.，1996；Dow et al.，1995；Schlotterer，2001），因而这类标记在基因组中具有最高的位点杂合度。标记在不同物种的图谱间是否可以转移取决于引物序列的保守性及其揭示的基因位点是否杂合，所以 SSR 在不同物种的图谱构建中具有最高的可转移性，因而是比较基因组研究的最有效的标记技术。利用设计的微卫星引物进行的作图研究结果显示，约为 51.2%的从序列直接设计的 SSR 可定位于所用家系的图谱上（Yin et al.，2004c），所以如果从已定位的标记中选择标记并用于图谱构建，标记将具有更高的转移效率。根据杨属树种已建成的微卫星图谱，我们可以定位选择部分标记，利用选出的标记可以在其他杨属树种和家系中快速经济地建成图谱，这样在一种杨属树种中获得的基因组信息立即可应用于其他的杨属树种，同时可以将其他杨属树种的基因组信息通过共有的垂直位点联系到共祖先图谱上，这样就可以对不同的研究所获得的信息进行比对，从而发现新的等位基因或者对不同家系所定位的 QTL 进行验证，利用微卫星图谱，我们可以将杨属的不同树种作为具有一套共祖先基因组（consensus genome）的大遗传系统进行研究。Yin 等（2004b）利用建成的图谱完成了对杨属树种已发表的 QTL 在共祖先图谱上的定位，充分说明了这一方法的有效性。在对杨属树种种间杂种的作图研究中，Yin 等（2001）发现第 4 和第 19 条染色体大部分区域上发生了偏分离，由于这两条染色体上定位了抗病基因，因此该文的作者提出，由于适应性基因在这两条染色体上积累，在选择作用下，可能会导致一种‘滚雪球’效应，从而使染色体分化在某条或某几条染色体上渐渐扩大，进而导致种的分化，这种分化机制不需要地理隔离也不需要染色体结构差异的前提。大量比较基因组研究表明，在生物的进化过程中，每条染色体的进化速率是不一样的。这种机制是否是种分化的原始机制还需要更有力的证据。例如在这一假说的前提下，适应性基因富集的染色体将会比其他染色体在更广泛的范围内检测到连锁不平衡。利用已定位的 SSR 标记，我们可以对杨属不同树种的基因组进行精细的比较研究，从而建立起染色体的重排随遗传距离变化的动态过程，这对于更深入地了解种分化的遗传机制有重要意义。这是将来杨属树种比较基因组研究的一个重要方向。

杨属树种是目前图谱信息最为丰富的树种，同时杨属树种的全基因组序列组装也即将完成，根据定位的微卫星引物序列，我们可以对相应的微卫星进行物理定位，这样如果

发现与目标性状连锁的微卫星标记,就可以将研究的基因组范围缩小到一个特定的目标区域,根据目标区域的序列,我们可以通过与其他物种如拟南芥 *Arabidopsis thaliana*(L.) Heynh.等的基因组进行比对,从而确定出可能的候选基因。例如,如果可以将目标 QTL 定位于 1 cM 的区间内,假定杨属树种基因组含有 5 万个功能基因,1 cM 的基因组区域平均含 20 个左右的基因,通过对候选基因进行研究可以很快找出目标基因。这一方案是后基因组时代功能基因研究中最有效、最基本的方法之一,随着杨属树种全基因组序列的组装完成,这一方法在杨属树种的功能基因组研究中将发挥重要作用,这也是未来杨属树种基因组研究的一个重要方向。

## 2.2 基因组测序及物理图谱

美国能源部实施的对毛果杨无性系'Nisqually-1'进行的全基因组测序计划已于 2003 年 12 月完成(Wullschleger et al., 2002a),全序列用鸟枪法测定,序列库中共含有 7649993 个序列片段,平均的 Q20 读序长度超过 625 bp。去除叶绿体基因组的污染,测得的序列大约为 8× 基因组长度。根据最新的序列信息推测,杨属树种的基因组比原来估计的 550 Mb 要小一些,最新物理长度估计区间为 480 – 520 Mb,所以目前在估算各种参数时,杨属树种基因组物理长度以取 520 Mb 为准。对杨属树种基因组序列的初步组装也已完成,初步组装的结果,共获得 6900 个序列骨架(sequence scaffold),这些骨架的序列总长约为基因组长度的 5.4 倍,覆盖的区域约为 465 Mb, 基本上覆盖了常染色质的大部分。拼接成的序列骨架的大小在 2 kb 到 6.72 Mb 之间,骨架大小对应骨架数的分布曲线峰值约在 8.2 kb。因此从序列直接拼接得到的染色体片段的数目远超过杨属树种 19 条染色单体的数目。遗传图谱是把不连续序列骨架连接为染色体片段的最有效的工具,为了对基因组序列进行正确组装,美国橡树岭国家实验室构建了一张覆盖全基因组的微卫星图谱。根据遗传图谱上定位的 SSR 标记的引物配对序列对初步组装的序列骨架进行了核查,结果发现大约 50%的骨架存在序列拼接错误。这是由于基因组序列中广泛存在的随机重复序列导致的。根据在微卫星图谱基础上进行的最新拼接,已拼接了 286 Mb 的无疑议序列(Yin et al., 2004c)。为尽量消除重复序列的影响,另外的 2.6 倍基因组的序列正在测序中。预期杨属树种基因组全序列的测定及拼接将于今年夏天完成(私人交流,G. A. Tuskan)。

除了上述的全基因组信息,瑞典、法国、加拿大和美国都进行了杨属树种大规模的 EST 测序项目,EST 测序研究进展以瑞典的研究为代表。迄今,国际上已完成了超过 20 万的 EST 序列读数。对杨属树种全基因组序列的注释(annotation)工作也于 2004 年初在美国能源部的两个实验室展开(私人交流, G. A. Tuskan)。在物理图谱构建方面,加拿大的基因组中心构建了约为 10 倍基因组长度的共含 45511 个 BAC 克隆的文库,文库是利用 *Hin*dIII 限制性内切酶对 Nisqually-1 的基因组进行不完全酶切的产物构建的。加拿大基因组中心对所有的 BAC 库末端的序列测定已完成,并初步拼接成了 4625 个染色体片段(私人交流,M. Marra, BC Genome Sciences Center),美国能源部橡树岭国家实验室在其中大约 225 Mb 的片段中设计了 SSR 引物(Yin et al., 2004c)。综合上述信息,在接合序列骨架、BAC 连续克隆群(BAC contig)及遗传图的基础上,有望完成序列图、物理图按染色体的组装,并在 3 张图之间进行相互检验,从而提高组装的准确性。

对比其他树种，杨属树种在基因组研究较深入的植物中处于较理想的系统学位置（图 1），因而便于开展与这些物种基因组的比较分析（Brunner et al.，2004）。通过与模式植物拟南芥的基因比较，可以发现候选的功能基因或对基因的功能进行推测，这一策略已成为在其他植物中寻找基因的基本的生物信息学手段。随着杨属树种全基因组测序的完成，杨属树种也将成为研究其他植物基因组的一个有价值的参考体系，特别是对多年生及木本植物的基因组研究有重要的参考价值。

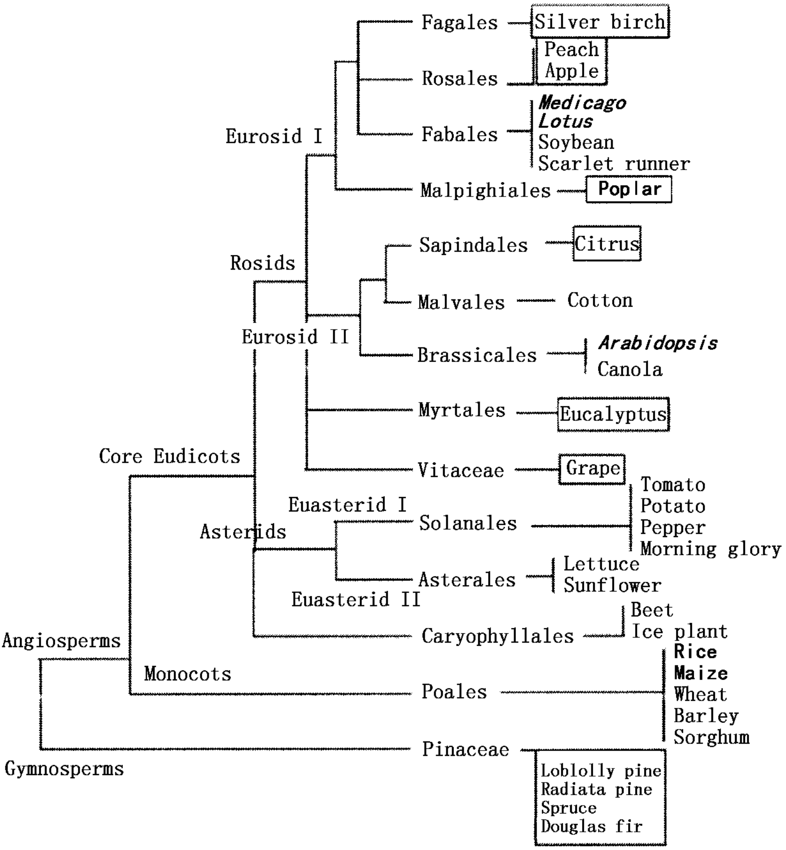


图 1　杨属树种与其他基因组研究较深入的植物的系统学关系比较　此图由 Brunner 等（2004）根据 Soltis 等（2000）的研究结果编辑。图中所列植物的基因组均测定了 1 万个以上的 EST 序列，物种名称黑体的为已经进行全序列测定或正在进行全序列测定的物种。物种名称未加边框为一年生草本，名称加框的则为木本植物。木本植物中共有 11 个物种已测定了 1 万个以上的 EST 序列，而杨属树种是目前唯一测定全序列的树种。相对于有全序列测定计划的其他植物如苜蓿、莲花、拟南芥、水稻和玉米，杨属树种具有利用这些植物的基因组信息开展比较与功能基因组研究较理想的系统学位置。

**Fig. 1.**　The phylogeny relationship between poplar and other plants with abundance genome information available. This figure is modified by Brunner et al.（2004）based on Soltis et al.（2000），showing only phylogeny orders that include genera with > 10000 expressed sequence tags（ESTs）. Species for which complete genome sequences are，or soon will be，available are in bold type；woody plants are in framework boxes and herbaceous plants without framework. There are 11 woody plants with more than 10000 ESTs sequenced and *Populus* is the only woody plants with complete genome sequences available. Compare to other species with complete genome sequences，such as medicago，lotus，arabidopsis，rice and maize，poplar is in a good phylogeny position to carry out comparative and functional genomics studies utilizing the genomics information from the above species.

## 2.3　基因芯片（microarray）

基因芯片分析是现代基因组研究的主要技术之一，广泛应用于人类、植物、动物及原核生物基因组研究中。基因芯片分析技术可分为两大类：一类是寡核苷酸芯片，另一类是

cDNA 芯片。这两类芯片在基因组研究中都有广泛的应用,例如在拟南芥中报道的利用寡核苷酸芯片进行的遗传图谱的快速构建(Cho et al.,1999)。而 cDNA 芯片是用于分析全基因组水平上基因表达的重要技术,利用 cDNA 芯片可以获得不同环境因子胁迫下的转录档案(transcript profiling),这是现代植物生理学的重要研究手段。随着杨属树种大规模 EST 计划的开展,cDNA 芯片在杨属树种的遗传及生理学研究中也开始得到应用,瑞典农业大学已开始利用 cDNA 芯片研究木材的发育过程,包括在细胞分裂、生长、次生细胞壁形成、木素形成及细胞死亡过程中的基因表达。初期建成的 cDNA 芯片利用了 2995 个 EST 序列,最近建成的 cDNA 芯片则包含了 13000 个以上的 EST 序列(Wullschleger et al.,2002a)。此外英国和美国利用 cDNA 芯片研究碳素代谢生理过程的项目也已经展开(私人交流,G. A. Tuskan)。虽然这种技术的高效性已被广泛接纳,然而目前还不是一项完全成熟的技术(Brunner et al.,2004),除了在实验设计、数据处理及实验技术方面存在的问题外,取样的时机对分析的影响很大,因为 mRNA 寿命较短,在原核生物中为 1–2 h,在高等植物中也只有 10–12 h,所以要获得理想的结果,需要的取样强度很大,某一瞬时的转录档案很难获得较好的代表性。这样就导致实验成本过高。另外由于表达量有差异的 mRNA 往往很多,后期筛选的工作量还是很大。还有仅根据 mRNA 转录水平的差异,我们还不能确定对应基因在表型上的效应。最近的研究已经开始将 cDNA 芯片上的差异表达与数量性状的分析相结合,美国北卡罗来纳州立大学 R. Sederoff 领导的研究小组在桉树 *Eucalyptus grandis* W. Hill ex Maiden 中率先将 cDNA 芯片上表达的数据作为数量性状进行研究,并在分离群体中定位(Wullschleger & Difazio,2003)。Kirst 等(2003)利用这一方法在桉树中定位到了两个较大效应的基因。由于转录水平上的差异表达可以通过编码区基因的多态性分离分析(segregation analysis)在群体中定位,这样就有可能区分影响转录水平的反式(*trans*,基因表达产物对转录水平的调控)或顺式(*cis*,基因结构序列对转录水平的调控)作用因子(Wullschleger et al.,2002b)。最近 Purdue 大学的 Doerge(2003)报道了将基于遗传图谱的 QTL 分析与 cDNA 芯片上检测到的表达水平的差异相结合的线性分析模型。通过两者的结合,有可能直接找到影响数量性状的基因或调控因子。由于杨属树种具有更完善的基因组信息,利用这一策略,美国北卡罗来纳州立大学与能源部橡树岭国家实验室正在联合开展杨属树种中影响碳固定及光合产物分配的 QTL 的研究(私人交流,G. A. Tuskan)。

### 2.4 连锁不平衡作图

基于标记和目的基因在家系(pedigree)子代中的分离,主要是分析标记与基因间的连锁关系(linkage)。而基于自然群体中的连锁不平衡(linkage disequilibrium,LD)分析则是研究标记与目标性状的相关(association)。两者的不同主要在于,连锁分析利用的是最近一代或几代的重组事件,而连锁不平衡分析则是利用群体历史上的重组事件。在历史上发生的重组使连锁的标记渐渐分布到不同的同源染色体上,这样就只有相隔很近的标记才能不被重组掉,从而形成大小不同的单倍型片段(haplotype block)。除了相隔很近的基因,某些相隔较远的基因由于受相同的选择压力,也可能产生连锁不平衡。连锁不平衡作图已成为人类疾病基因原位克隆(positional cloning)的主要手段之一(Hall et al.,2002),并在动植物复杂性状的研究中也得到广泛应用。特别是人类基因组研究中,连锁不平衡作图

是继全基因组测序项目之后的最大的研究项目。人类和模式植物的 LD 研究显示，在不同基因组区域，LD 的分布和延伸范围都有很大差异（Hall et al.，2002；Nordborg et al.，2002）。通过双位点标记的模拟分析，Kruglyak（1999）得出的结论显示在人基因组中可应用的 LD 一般不超过 3 kb。木本植物大多为自由授粉的异交物种，一般认为木本植物天然群体中的 LD 程度很低，这样木本植物可利用的 LD 可能会仅局限于非常小的区域，这是在木本植物中应用分子标记辅助选择育种的主要障碍。同时这也意味着在木本植物中进行 LD 作图将需要大量的标记。所以理论上 LD 分析在木本植物中的应用效率是很低的，从这一点上看，基于家系的连锁分析仍将是 QTL 分析的主要手段（Yin et al.，2004a）。但是 LD 延伸范围的大小是一把双刃剑，如果 LD 延伸范围很大，则容易发现相关标记，然而位于同一单倍型片段上的标记将不能为近一步靠近目的基因提供信息。如果 LD 延伸范围很小，则较难发现相关标记，但一旦发现相关标记就会非常靠近目标基因，这对于从成簇分布的同源基因（如抗病基因，R-gene）中克隆目的基因尤其有效。在连锁图谱部分的讨论中，我们提到可以利用连锁图谱结合目标区域的序列信息确定候选基因（candidate gene），然后就可以根据候选基因的序列设计引物进行 LD 分析，这是目的基因原位克隆的一种最有效的方法。由于很快将有杨属树种全基因组序列信息，连锁图谱与 LD 分析相结合的方法将是在杨属树种中克隆目的基因的有效方法。

迄今，我们对 LD 在木本植物基因组中分布的了解都是基于理论上的推测，并没有实验的数据支持。以往研究进行的 LD 分析都是基于随机取样的标记，由于标记间的遗传距离及物理距离不清楚，无法对 LD 的分布模式进行研究。最近已有几个研究开始对针叶树基因组中的 LD 分布进行探讨，但这些研究仅是基于遗传距离已知的标记，标记间的物理距离还无从得知（Wilcox et al.，2004）。Yin 等（2004a）在杨属树种中对抗锈病基因座的附近区域进行了木本植物中局部区域 LD 的最为细致的研究。根据遗传图谱，Yin 等排出了 *MXC3* 附近一个 40 kb 的区域，这样就获得了该区域内标记的物理距离。在克隆 *MXC3* 基因的过程中，Stirling 等（2001）发现了该基因座位附近存在强烈的重组抑制，因而未能成功克隆该基因。如果该区域重组被抑制，理论上该区域会成为存在强烈连锁不平衡的单倍型片段，LD 分析的结果显示该区域被分成了 3 个小的单倍型片段，因而否定了 Stirling 等（2001）的推测。由于该基因座位被定位于其中一个连锁群的末端，因此，Yin 等（2004a）推测，Stirling 等（2001）的发现可能是由于作图亲本中对应的同源染色单体长度有差异（该亲本为种间杂种），从而在该染色体对末端的不平衡区域内没有重组发生。这样 Stirling 等（2001）对该基因克隆失败的主要原因可能是由于作图亲本的同源染色体结构差异所致，Yin 等（2004a）在另一个作图群体中发现该区域有正常的重组频率。通过这两个研究的比较，我们能够看出 LD 分析可以提供目标基因区更详细的信息。同时从上面研究我们还可以看出，如果在很小的区域内同时存在两个以上的同源基因，基于家系的连锁分析是很难提供足够精细的分辨率对它们加以区分的，这时 LD 分析可能是最有效的选择。同时，在 *MXC3* 附近区域的 LD 分析中，根据标记的物理距离，Yin 等（2004a）发现，亚单倍型片段（sub haplotype block）间区域内的重组频率明显高于期望值，这与理论上假设单倍型片段间由重组热点（recombination hotspot）间隔的理论相一致（Hall et al.，2002）。由于杨属树种的全基因组序列已知，这样就可以在基因组中选择间隔不同物理长度的标记进

行 LD 作图,从而研究 LD 在杨属树种基因组中的分布。由于对木本植物群体中 LD 的了解是一个基本的理论问题并且是限制标记辅助选择在木本植物中应用的瓶颈,LD 作图是未来杨属树种基因组研究的一个重要方向。

## 3 杨属树种基因组在基本科学问题研究中的应用

除了上面提到的进化、比较及功能基因组研究,随着全基因组序列的测定,杨属树种作为一种模式植物也将在以下几个基本的科学问题的探索中发挥作用。

### 3.1 QTL 的遗传学本质

与经典的微效多基因模式不同,基于图谱上的 QTL 分析发现大多数数量性状都是由为数不多的几个主效基因控制的。迄今为止,QTL 分析所发现的控制某一数量性状的 QTL 数目一般在 10 个以下(Kearsey & Farquhar, 1998)。由于群体大小、分析方法等技术因素的限制, 虽然我们对 QTL 有了不同的认识,但目前仍不能推翻传统的假说(Barton & Keightley, 2002)。然而基于图谱的 QTL 分析的确使 QTL 变成了遗传上可操作的单位,最近从玉米 *Zea mays* L.和水稻 *Oryza sativa* L.中克隆的控制分蘖的基因及水稻中发现的矮化基因都在实验证据上强化了我们对于数量性状是受主效基因控制的印象(Barton & Keightley, 2002)。随着对转录过程的深入研究,我们对基于图谱研究发现的 QTL 有了另一个层面的认识,除了功能主效基因外,主效的 QTL 也可能是某一类的转录调控因子。转录调控因子可分成两大类:一类是序列调控因子,也称为顺式调控因子;另一类是转录或表达产物调控因子,也称为反式调控因子。某个顺式或反式调控因子可以特异地调控某个基因的转录,也可能调控某一类基因的转录,或者有更广泛的作用对象。在植物的进化过程中,形成了非常复杂的代谢途径,代谢途径中某一基因的突变往往并不会阻断正常的生理过程并产生表型上的效应。但是调控某一类基因的调控因子如果发生了突变,则可能导致表型上的变化,并呈现较大的效应。所以 QTL 分析中定位到的 QTL 并不一定是直接的目的基因,也可能是控制基因转录的调控因子。前面提到的利用基因芯片和遗传图相结合的方法,可以对这一问题进行研究。基于前面提到的类似的策略,最近克隆了控制番茄 *Lycopersicon esculentum* Mill.果实大小的 QTL——*fw2.2*。这个 QTL 的效应是由该基因座位内的一个单一基因 *ORFX* 导致的。根据该基因序列预测的蛋白与人类肿瘤基因 *RAS-P21* 表达的蛋白存在结构上的相似性。等位基因分析表明,*fw2.2* 座位在表型上的差异可能不是由于基因编码区的碱基差异所致,而是由于基因上游的启动子区域碱基变异所致(Frary et al., 2000)。基因调控区的变异也同样用于解释玉米驯化过程中控制分蘖的主效 QTL—— *teosinte branched1* 的效应(Wang et al., 1999)。然而这些假定的验证还需要更直接的证据。随着杨树基因组研究的深入,我们可能在 QTL 的遗传学本质的研究方面提供新的证据。

### 3.2 重组的机制

对原核生物中的重组机制及过程现在已经比较清楚,如大肠杆菌 *Escherichia coli* 的重组过程主要如下:重组复合物(RecA、RecB、RecC、RecD)识别 chi 位点, chi 位点为 8 个寡核苷酸的重组热点序列;RecA 可以促进同源序列的联会, RecB、RecC 及 RecD 则具有解旋酶和双链 DNA 内切酶活性,在这几个酶作用下会在解旋过程中降解 DNA 链 3′端,或者

在 chi 位点 3′端产生几个碱基的缺刻。然后 RuvA 蛋白形成一个四聚体从而提供一个同源 DNA 链交叉（Holliday junction）的平台，接着 2 个 RuvB 六聚体形成哑铃环状结构推动重组蛋白复合体沿 DNA 链移动。最后，同源 DNA 链交叉的暴露面与 RuvC 核酸酶结合，RuvC核酸酶在偏爱的切割序列 A/TTTG/C 处将 4 条单链两两切开。这一过程中已发现有不同的同工酶，但是基本作用过程是一样的。由于不同的链切割和重连结方式，产生不同的链分离方式，一种为切割（splice）结构，另一种为补丁（patch）结构，其中补丁结构的分离方式不产生链交换（仅在 RuvB 推动的区域内交换 1000 个左右的碱基），这种形式的重组主要有修复功能。而切割结构的分离方式则导致同源 DNA 链的互换。所以重组过程中是否发生链交换决定于同源 DNA 链交叉的分离方式。对于真核生物的重组过程还不是很清楚，但是在真核生物中已发现了大量 RecA 的同源物，然而同源 DNA 链交叉是否在 Ruv 蛋白类似物作用下分离还不清楚（Arnold & Kowalczykowski，2000；Amundsen et al.，2000）。利用 LD 分析进行重组热点作图是目前人类基因组研究的热点之一。真核生物中是否有特异的重组热点序列是一个基本的科学问题。另外，重组热点是否在染色体上随机分布，染色体不同区域检测到的重组频率的差异是否可能由于在某些区域有偏好的分离方式也是一些基本的科学问题。通过标记间的 LD 分析、遗传图距及物理图距的比对，结合特定区域内细胞学研究观察到的减数分裂过程中的同源染色体交叉数可以对上述问题进行研究。

### 3.3　多年生植物发育调控模式

木本植物不同于一年生的草本植物，木本植物除了有木质部这一特殊组织外，在木本植物的生长发育过程中还存在幼、成期的转换并随季节更替进行有规律的发育调控。同时木本植物的个体寿命很长，在生长过程中遇到的环境胁迫也更为复杂，因此木本植物可能有更为复杂的对付环境胁迫及病原物侵染的反应防御系统。到底是因为哪些特异的基因或基因家族导致了木本植物与草本植物的不同呢？这是杨属树种基因组研究需要回答的一个最有价值的科学问题。根据瑞典进行的约 20000 个 EST 在杨树和拟南芥中的比对，发现杨树特有的 EST 大约只有 20 个（私人交流，S. Jansson），由于这些 EST 未经组装，因而许多 EST 可能来源于同一基因，但从统计学的角度看，杨树中不同于拟南芥的特有基因（unigene）是很有限的，是否这些有限的特异基因创造了一种全新的生物体系呢？现在的观点认为这两个物种的差异可能更多的决定于它们有不同的基因发育调控和互作模式。基因的表达调控，除了转录因子的控制，染色体的包装方式的差异在大的系统上可能会有更重要的影响。如异染色质区域为基因表达不活跃区，这样同源基因在不同物种中由于包装方式不同会产生不同的表达水平和不同的调控模式。最近提出的‘组蛋白密码（histone code）’假说可能会成为生物学研究史上的另一个重要的里程碑（Jenuwein & Allis，2001）。这一假说的思想大致如下：虽然组蛋白在整个生物界都是相当保守的，但组蛋白的不同修饰方式对基因的表达有决定作用，一般组蛋白的乙酰化和磷酸化会产生松散包装，从而有利于基因的转录和表达，组蛋白的甲基化则产生紧密包装，而不利于基因的转录和表达。一个被包装的基因是否处于活跃状态，则最终取决于组蛋白不同位置上的乙酰化、磷酸化和甲基化的组合，而这种组合形成了决定基因是否可以表达的‘组蛋白密码”，这是继遗传密码之后的另一套密码。这套密码决定基因的表达和沉默，因此这一学

说具有重大的科学意义。自这一学说提出后的短时间内,该领域的研究已成为现代遗传学研究最热门的一个领域。目前组蛋白密码还只是一种学说,对组蛋白密码的破译研究才刚刚开始,是否存在一种通用的组蛋白密码需要在不同的生物中开展研究。由于杨属树种具有较深的基因组学研究基础,如果我们能够找出其常染色质和异染色质区的同源基因,从而研究包装这些基因的组蛋白在不同位置的不同修饰方式,将对'组蛋白密码'的破译发挥重要作用。根据最近的细胞学研究,杨属树种的硕大染色体的一条臂是由异染色质组成的(Yin et al.,未发表),在哺乳动物中发现,一些新物种的产生是由于染色体上新出现的异染色质臂导致(Gibson, 1984)。因此结合序列及细胞学的研究,将为上述问题的研究提供一些有效的线索。

除了转录因子和组蛋白包装,基因的互作也是一个重要的方面,例如在对病原菌的防御体系中,一般认为,特异的垂直抗性基因对其他病原菌也存在水平抗性,并且通过基因间的上位性互作形成一个复杂的防御网络。特异抗性机制现在已研究得比较清楚,但非特异的抗性机制我们还所知有限,上位性互作是一个重要的研究手段。除了在防御体系研究中的意义,基因的上位性效应也是决定发育过程的重要遗传学因素,对上位性效应的深入研究将有助于我们对多年生植物发育调控机制的了解,这是近年来基因组学研究领域的一个重点并将在未来研究中占有越来越重要的位置。

## 参 考 文 献

Abecasis G R, Cherny S S, Cardon L R. 2001. The impact of genotyping error on family-based analysis of quantitative traits. European Journal of Human Genetics 9: 130 – 134.

Altukhov I, Salmenkova E A. 2002. DNA polymorphism in population genetics. Genetica 38: 1173 – 1195.

Amundsen S K, Taylor A F, Smith G R. 2000. The RecD subunit of the *Escherichia coli* RecBCD enzyme inhibits RecA loading, homologous recombination, and DNA repair. Proceedings of the National Academy of Sciences USA 97: 7399 – 7404.

Anh S, Tanksley S D. 1993. Comparative linkage maps of the rice and maize genomes. Proceedings of the National Academy of Sciences USA 90: 7980 – 7984.

Arnold D A, Kowalczykowski S C. 2000. Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme. Journal of Biological Chemistry 275: 12261 – 12265.

Barreneche T, Bodenes C, Lexer C, Trontin J F, Fluch S, Streiff R, Plomion C, Rousse G L, Steinkellner H, Burg K, Favre J M, Glossl J, Kremer A. 1998. A genetic linkage map of *Quercus robur* L. (Pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. Theoretical and Applied Genetics 97: 1090 – 1103.

Barton N H, Keightley P D. 2002. Understanding quantitative genetic variation. Nature Reviews Genetics 3: 11 – 21.

Binelli G, Bucci G. 1994. A genetic linkage map of *Picea abies* Karst, based on RAPD markers, as a tool in population genetics. Theoretical and Applied Genetics 88: 283 – 288.

Blackburn K B. 1924. A preliminary account of the chromsomes and chromosome behavior in the Salicaceae. Annals of Botany 38: 361 – 378.

Bradshaw H D, Ceulemans R, Davis J, Stettler R. 2000. Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. Journal of Plant Growth Regulation 19: 306 – 313.

Bradshaw H D, Villar M, Watson B D, Otto K G, Stewart S, Stettler R F. 1994. Molecular genetics of growth and development in *Populus* III: a genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theoretical and Applied Genetics 89: 167 – 178.

Brunner A M, Busov V B, Strauss S H. 2004. Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends in Plant Science 9: 49 – 56.

Byrne M, Marquezgarcia M I, Uren T, Smith D S, Moran G F. 1996. Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. Australian Journal of Botany 44: 331 – 341.

Cevera M T, Storme V, Ivens B, Gusmao J H, Liu B, Hostyn V, Slycken J V, Van Montagu M, Boerjan W. 2001. Dense genetic linkage maps of three *Populus* species (*P. deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. Genetics 158: 787 – 809.

Cho R J, Mindrinos M, Richards D R, Sapolsky R J, Anderson M, Drenkard E, Dewdney J, Reuber T L, Stammers M, Federspiel N, Theologis A, Yang W H, Hubbell E, Au M, Chung E Y, Lashkari D, Lemieux B, Dean C, Lipshutz R J, Ausubel F M, Davis R W, Oefner P J. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nature Genetics 23: 203 – 207.

Costa P, Pot D, Dubos C, Frigerio J M, Pionneau C, Bodenes C, Bertocchi E, Cervera M T, Remington D L, Plomion C. 2000. A genetic map of maritime pine based on AFLP, RAPD and protein markers. Theoretical and Applied Genetics 100: 39 – 48.

Deorge R W. 2003. Functional genomics of quantitative traits: old methods for new ideas. Plant and Animal Genome Conference XII, 10 – 14 January 2004, San Diego, CA, USA. 68.

Devey M E, Bell J C, Smith D N, Neale D B, Moran G F. 1996. A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers. Theoretical and Applied Genetics 92: 673 – 679.

Devos K M, Gale M D. 1997. Comparative genetics in the grasses. Plant Molecular Biology 35: 3 – 15.

Dow B D, Ashley M V, Howe H F. 1995. Characterization of highly variable (GA/CT)n microsatellites in the bur oak, *Quercus macrocarpa*. Theoretical and Applied Genetics 91: 137 – 141.

Echt C S, Nelson C D. 1997. Linkage mapping and genome length in eastern white pine (*Pinus strobus* L.). Theoretical and Applied Genetics 94: 1031 – 1037.

Eckenwalder J E, Bradshaw H D, Heilman P E, Hinckley T M. 1996. Systematics and evolution of *Populus*. In: Stettler R F ed. Biology of *Populus* and Its Implications for Management and Conservation. Ottawa: National Research Council of Canada. 7 – 32.

Frary A, Nesbitt T C, Frary A, Grandillo S, Knaap E V D, Cong B, Liu J, Meller J, Elber R, Alpert K B, Tanksley S D. 2000. fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85 – 88.

Gibson L J. 1984. Chromosomal changes in Mammalian speciation: a literature review. Origins 11: 67 – 89.

Graf J. 1921. Beitrage zur Kenntnis der Gattung *Populus*. Berichte der Deutschen Botanischen Gesellschaft 38: 405 – 434.

Grattapaglia D, Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. Genetics 137: 1121 – 1137.

Hall D, Wijsman E M, Roos J L, Gogos J A, Karayiorgou M. 2002. Extended intermarker linkage disequilibrium in the Afrikaners. Genome Research 12: 956 – 961.

Harrison J W H. 1924. A preliminary account of the chromosomes and chromosome behavior in the Salicaceae. Annals of Botany 38: 361 – 378.

Jenuwein T, Allis C D. 2001. Translating the histone code. Science 293: 1074 – 1080.

Johnson H. 1940. Cytological studies of diploid and triploid *Populus tremula* and of crosses between them. Hereditas 26: 321 – 352.

Kearsey M J, Farquhar A G L. 1998. QTL analysis in plants; where are we now? Heredity 80: 137 – 142.

Kirst M, Myburg A A, Sederoff R R. 2003. Genetical genomics of *Eucalyptus*: combining expression profiling and genetic segregation analysis. Plant and Animal Genome Conference XI, 11 – 15 January 2003, San Diego, CA, USA. 36.

Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genetics 22: 139 – 144.

Lagercrantz U. 1998. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent

rearrangements. Genetics 150: 1217 – 1228.
Liu Z, Furnier G R. 1993. Inheritance and linkage of allozymes and restriction fragment length polymorphisms in trembling aspen. Journal of Heredity 84: 419 – 424.
Mayr E. 1970. Populations, species, and evolution. Cambridge: Belknap Press.
Moore G, Devos K M, Wang Z, Gale M. 1995. Cereal genome evolution. Grasses, line up and form a circle. Current Biology 5: 737 – 739.
Mukai Y, Suyama Y, Tsumura Y, Kawahara T, Yoshimaru H, Kondo T, Tomaru N, Kuramoto N, Murai M. 1995. A linkage map for sugi (*Cryptomeria japonica*) based on RFLP, RAPD, and isozyme loci. Theoretical and Applied Genetics 90: 835 – 840.
Muntzing A. 1936. The chromosomes of a giant *Populus tremula*. Hereditas 21: 383 – 393.
Nelson C D, Kubisiak T L, Stine M, Nance W L. 1994. A genetic linkage map of longleaf pine (*Pinus palustris* Mill.) based on random amplified polymorphic DNAs. Journal of Heredity 85: 433 – 439.
Nelson C D, Nance W L, Doudrick R L. 1993. A partial genetic linkage map of slash pine (*Pinus elliottii* Engelm. var. *elliottii*) based on random amplified polymorphic DNAs. Theoretical and Applied Genetics 87: 145 – 151.
Nordborg M, Borevitz J O, Bergelson J, Berry C C, Chory J, Hagenblad J, Kreitman M, Maloof J N, Noyes T, Oefner P J, Stahl E A, Weigel D. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 30: 190 – 193.
Paterson A H, Bowers J E, Burow M D, Draye X, Elsik C G, Jiang C X, Katsar C S, Lan T H, Lin Y R, Ming R G, Wright R J. 2000. Comparative genomics of plant chromosomes. The Plant Cell 12: 1523 – 1540.
Remington D L, Wheten R W, Liu B H, O'Malley D M. 1999. Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. Theoretical and Applied Genetics 98: 1279 – 1292.
Schlotterer C. 2001. Genealogical inference of closely related species based on microsatellites. Genetics Research 78: 209 – 212.
Smith E C. 1943. A study of cytology and speciation in the genus *Populus* L. Journal of the Arnold Arboretum 24: 275 – 305.
Soltis D E, Soltis P S, Chase M W, Mort M E, Albach T D, Zanis M, Nixon K C, Hahn W H, Hoot S B, Fay M F, Prince L M, Kress W J, Farris J S. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. Botanical Journal of the Linnean Society 133: 381 – 461.
Stirling B, Newcombe G, Vrebalov J, Bosdet I, Bradshaw H D. 2001. Suppressed recombination around the MXC3 locus, a major gene for resistance to poplar leaf rust. Theoretical and Applied Genetics 103: 1129 – 1137.
Strauss S, Lande R, Namkoong G. 1992. Limitations of molecular-marker-aided selection in forest tree breeding. Canadian Journal of Forest Research 22: 1050 – 1061.
Taylor G. 2002. *Populus*: *Arabidopsis* for forestry. Do we need a model tree? Annals of Botany 90: 681 – 689.
Temesgen B, Brown G R, Harry D E, Kinlaw C S, Sewell M M, Neale D B. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theoretical and Applied Genetics 102: 664 – 675.
Travis S E, Ritland K, Whitham T G, Keim P. 1998. A genetic linkage map of Pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. Theoretical and Applied Genetics 97: 871 – 880.
Tsuchiya T. 1984. Problems in linkage mapping in barley. Barley Genetics Newsletter 14: 85 – 88.
Tuskan G A, Gunter L E, Yang Z M, Yin T M, Sewell M M, Difazio S P. 2004. Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa*. Canadian Journal of Forest Research 34: 85 – 93.
Van Dillewijn C. 1940. Zytologische Studien in der Gattung *Populus*. Genetica 22: 131 – 182.
Viruel M A, Messeguer R, Devicente M C, Garciamas J, Puigdomenech P, Vargas F, Arus P. 1995. A linkage map with RFLP and isozyme markers for almond. Theoretical and Applied Genetics 91: 964 – 971.
Wang C (**王战**), Tung S-L (**董世林**), Yang C-Y (**杨昌友**). 1984. *Populus* L. In: Flora Reipublicae Popularis Sinicae (**中国植物志**). Beijing: Science Press. 20 (2): 1 – 94.
Wang R L, Stec A, Hey J, Lukens L, Doebley J. 1999. The limits of selection during maize domestication. Nature 398: 236 – 239.
White M J D. 1978. Modes of speciation. San Fransisco: W H Freeman and Company.

Whitkus R, Doebley J, Lee M. 1992. Comparative genome mapping of sorghum and maize. Genetics. 132: 1119 – 1130.

Wilcox P L, Cato S A, McMillan L K, Power M B, Burdon R D, Echt C S. 2004. Patterns of linkage disequilibrium in *Pinus radiata*. Plant and Animal Genome Conference XII, 10 – 14 January 2004, San Diego, CA, USA. 68.

Wu R L, Han Y F, Hu J J, Fang J J, Li L, Li M L, Zeng Z B. 2000. An integrated genetic map of *Populus deltoides* based on amplified fragment length polymorphisms. Theoretical and Applied Genetics 100: 1249 – 1256.

Wullschleger S D, Difazio S P. 2003. Emerging use of gene expression microarrays in plant physiology. Comparative and Functional Genomics 4: 216 – 224.

Wullschleger S D, Jansson S, Taylor G. 2002a. Genomics and forest biology: *Populus* emerges as the perennial favortie. The Plant Cell 14: 2651 – 2655.

Wullschleger S D, Tuskan G A, Difazio S P. 2002b. Genomics and the tree physiologist. Tree Physiology 22: 1273 – 1276.

Yin T M, Difazio S P, Gunter L E, Jordy S, Tuskan G A. 2004a. Mapping the rust resistant loci MXC3 and MER in *P. trichocarpa* and assessing the intermarker linkage disequilibrium in MXC3 region. New Phytologist. (in press with online doi: 10.1007/500122-004-1635-5)

Yin T M, Difazio S P, Gunter L E, Riemenschneider D, Tuskan G A. 2004b. Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. Theoretical and Applied Genetics. (in press with online doi: 10.1111/j01469-8137.2004.01161.x)

Yin T M, Huang M R, Wang M X, Zhu L H, Zeng Z B, Wu R L. 2001. Preliminary interspecific genetic maps of the *Populus* genome constructed from RAPD markers. Genome 4: 602 – 609.

Yin T M, Wang X R, Andersson B, Lerceteau-Kohler E. 2003. Nearly complete genetic maps of *Pinus sylvestris* L. (Scots pine) constructed by AFLP marker analysis in a full-sib family. Theoretical and Applied Genetics 106: 1075 – 1083.

Yin T M, Zhang X Y, Huang M R, Wang M X, Zhuge Q, Tu S M, Zhu L H, Zeng Z B, Wu R L. 2002. Molecular linkage maps of *Populus* genome. Genome 45: 541 – 555.

Zamir D T Y. 1986. Unequal segregation of nuclear genes in plants. Botanical Gazette 147: 355 – 358.

Zhang D, Zhang Z, Yang K, Li B. 2004. Genetic mapping in (*Populus tomentosa* × *Populus bolleana*) and *P. tomentosa* Carr. using AFLP markers. Theoretical and Applied Genetics 108: 657 – 662.